Informative Feature Selection for Object Recognition via Sparse PCA

Nikhil Naikal
Allen Yang
S. Shankar Sastry

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2011-27
http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-27.html

April 7, 2011

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
07 APR 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE
Informative Feature Selection for Object Recognition via Sparse PCA

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Berkeley,Department of Electrical Engineering and Computer Science,Berkeley,CA,94720

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Bag-of-words (BoW) methods are a popular class of object recognition methods that use image features (e.g. SIFT) to form visual dictionaries and subsequent histogram vectors to represent object images in the recognition process. The accuracy of the BoW classifiers, however, is often limited by the presence of uninformative features extracted from the background or irrelevant image segments. Most existing solutions to prune out uninformative features rely on enforcing pairwise epipolar geometry via an expensive structure-from-motion (SfM) procedure. Such solutions are known to break down easily when the camera transformation is large or when the features are extracted from lowresolution low-quality images. In this paper, we propose a novel method to select informative object features using a more efficient algorithm called Sparse PCA. First, we show that using a large-scale multiple-view object database, informative features can be reliably identified from a highdimensional visual dictionary by applying Sparse PCA on the histograms of each object category. Our experiment shows that the new algorithm improves recognition accuracy compared to the traditional BoW methods and SfM methods. Second, we present a new solution to Sparse PCA as a semidefinite programming problem using Augmented Lagrange Multiplier methods. The new solver outperforms the state of the art for estimating sparse principal vectors as a basis for a low-dimensional subspace model. The source code of our algorithms will be made public on our website.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**10**	

Copyright © 2011, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Informative Feature Selection for Object Recognition via Sparse PCA *

Nikhil Naikal, Allen Yang and Shankar Sastry
EECS Dept. at UC Berkeley
Berkeley, California.
{nnaikal, yang, sastry}@eecs.berkeley.edu

Abstract

*Bag-of-words (BoW) methods are a popular class of object recognition methods that use image features (*e.g.*, SIFT) to form visual dictionaries and subsequent histogram vectors to represent object images in the recognition process. The accuracy of the BoW classifiers, however, is often limited by the presence of uninformative features extracted from the background or irrelevant image segments. Most existing solutions to prune out uninformative features rely on enforcing pairwise epipolar geometry via an expensive structure-from-motion (SfM) procedure. Such solutions are known to break down easily when the camera transformation is large or when the features are extracted from low-resolution, low-quality images. In this paper, we propose a novel method to select informative object features using a more efficient algorithm called Sparse PCA. First, we show that using a large-scale multiple-view object database, informative features can be reliably identified from a high-dimensional visual dictionary by applying Sparse PCA on the histograms of each object category. Our experiment shows that the new algorithm improves recognition accuracy compared to the traditional BoW methods and SfM methods. Second, we present a new solution to Sparse PCA as a semidefinite programming problem using Augmented Lagrange Multiplier methods. The new solver outperforms the state of the art for estimating sparse principal vectors as a basis for a low-dimensional subspace model. The source code of our algorithms will be made public on our website.*

1. Introduction

In the past decade, the exponential growth of storage capacity has encouraged people to upload personal images to large online image databases such as Picassa and Flickr. The proliferation of modern smartphones equipped with low-quality mobile cameras has also garnered interest to endow smartphone users with the ability to automatically recognize common objects and landmark buildings in man-made urban environments. The existence of common objects and landmarks in these images has motivated research in visual object recognition [7, 9, 12, 27]. Images in these coarsely labelled databases are used to train classifiers that can be used to recognize different object categories. To tackle the problem of recognizing a large number of objects in large image databases, a visual-dictionary based approach has been well studied [19, 21], which have further led to several other methods to recognize objects in both the single-view and multi-view settings [3, 4, 8, 17, 23, 25]. Essentially, most of the methods work with certain visual descriptors (*e.g.*, SIFT and its many variants) extracted from the images to construct visual histograms, which represent the object appearance in the images using a precomputed visual dictionary.

Although the visual-dictionary methods have proven to be efficient in describing object images, the accuracy of the classifiers is often limited by the presence of uninformative image features typically extracted from the background or irrelevant image segments, such as pedestrians and vegetation (see Figure 1 for an example). When the irrelevant segments take on a significant portion of an image, the uninformative features can dominate the representation in the visual histogram, and hence lead to inferior recognition accuracy. In [24], Turcot and Lowe suggested, if a subset of so-called *useful features* or *informative features* can be systematically selected during the training stage, it not only further reduces the number of visual descriptors needed, but also significantly improves the recognition accuracy. Since in man-made environments, most objects of interest, in particular landmark buildings, are rigid objects, 3-D perspective geometry can be leveraged to select informative features that satisfy a pairwise epipolar constraint via RANSAC. This is known as the Structure-from-Motion (SfM) approach.

*This work was supported in part by ARO MURI W911NF-06-1-0076 and ARL MAST-CTA W911NF-08-2-0004. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute for Government purposes notwithstanding any copyright notation hereon.

Motivated by the literature, in this paper, we study how to improve informative feature selection in both speed and accuracy from possibly *low-resolution*, *low-quality* camera networks. One major problem in enforcing the epipolar constraint on images collected from low-power camera networks instead of high-end photography is that establishing wide-baseline feature correspondence of SIFT-type features is known to be brittle even using state-of-the-art bundle adjustment techniques [22]. In addition, the quality of images sampled from low-power camera sensors also presents a challenge to reliably extract image features to describe the appearance of interesting objects in multiple views.

We propose to address this problem by a principled semidefinite programming (SDP) technique, known as Sparse Principal Components Analysis (Sparse PCA) [30]. As an extension of the popular PCA method, Sparse PCA addresses a drawback of classical PCA that the principal vectors (PVs) as a basis of a low-dimensional subspace typically have dense non-zero entries. In particular, in high-dimensionality setting, the dense linear combinations of all the variables make it difficult to interpret the corresponding principal components (PCs).

In case of visual-dictionary based object recognition, the variables in a high-dimensional histogram are associated with the codewords that represent either informative foreground features or uninformative background. We contend that in a large-scale object image database, the subset of informative features can be reliably selected by the sparse coefficients in the first few PVs. The new solution is more robust to wide-baseline camera transformation and numerically more efficient than the existing solutions of establishing pairwise rigid-body correspondence.

1.1. Main Contributions

In this paper, we exploit the use of Sparse PCA as a variable selection tool for selecting informative features in the object images captured from low-resolution camera sensor networks. Firstly, we present a scheme for using Sparse PCA with high-dimensional covariance matrices constructed from visual histograms to extract a sparse support of visual codewords for each object category. We compare its performance with the SfM technique applied to large-baseline, low-quality multiple-view images. Secondly, we propose a state-of-the-art algorithm to speed up Sparse PCA using the Augmented Lagrange Multiplier (ALM) approach [2, 26]. To mitigate the high dimensionality of the visual dictionary, a direct variable elimination method called SAFE is presented to further prune out uninformative features for object recognition prior to the Sparse PCA process. The experiment on synthetic data shows that the new algorithm outperforms the previous convex programming algorithm (DSPCA) [5] in terms of speed while maintaining the same estimation accuracy. Finally, we perform object recognition experiments, which demonstrate improved recognition by successfully suppressing uninformative features. To aid peer evaluation, the source code of our algorithms will be made public on our website.

2. Recognition via Vocabulary Trees

In object recognition, certain local invariant features have become a popular representation of the object images, which can be extracted and encoded into high-dimensional descriptors using algorithms such as SIFT [15] and SURF [1]. In the bag-of-words (BoW) approach, these invariant features are further quantized to form a dictionary of *visual words*. All the feature descriptors in the training set are hierarchically clustered into visual word clusters (*e.g.*, using hierarchical *k*-means [13]). This hierarchical tree is commonly referred to as a *vocabulary tree* [19]. The size of a vocabulary tree for a large database ranges from thousands to hundreds of thousands. For example, in this paper, we use hierarchical *k*-means to construct 1,000-D vocabularies for our training image database, with a branch factor of k = 10 and four hierarchies.

To start the training process, feature descriptors in each training image are propagated down the vocabulary tree to form a BoW model for the image. Then a term-frequency inverse-document-frequency (*tf-idf*) weighted visual histogram $\boldsymbol{y}$ is defined for each training image [19]. For each object category, $i = 1 \cdots C$, m weighted histograms are generated from the m training images of that category respectively: $A_i = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_m\}$. All the C sets form the training set, $A = \{A_1, A_2, \cdots, A_C\}$.

During the testing phase, feature descriptors are extracted for the query image and propagated down the vocabulary tree by the same fashion to obtain a single weighted query histogram $\boldsymbol{q}$. Using the simplest nearest-neighbor classifier,[1] the query image is then given a relevance score s based on the ℓ_1-normalized difference between the weighted query and the ith training set A_i:

$$s(\boldsymbol{q}, A_i) = \min_{\boldsymbol{y}_j \in A_i} \| \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|_1} - \frac{\boldsymbol{y}_j}{\|\boldsymbol{y}_j\|_1} \|_1. \quad (1)$$

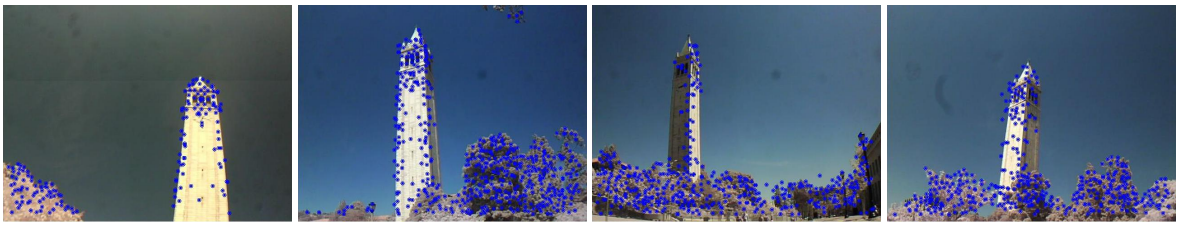
Finally, the label of the visual histogram $\boldsymbol{q}$ is assigned as the object category that achieves the minimal relevance score:

$$\text{label}(\boldsymbol{q}) = \arg\min_{i \in [1 \cdots C]} s(\boldsymbol{q}, A_i). \quad (2)$$

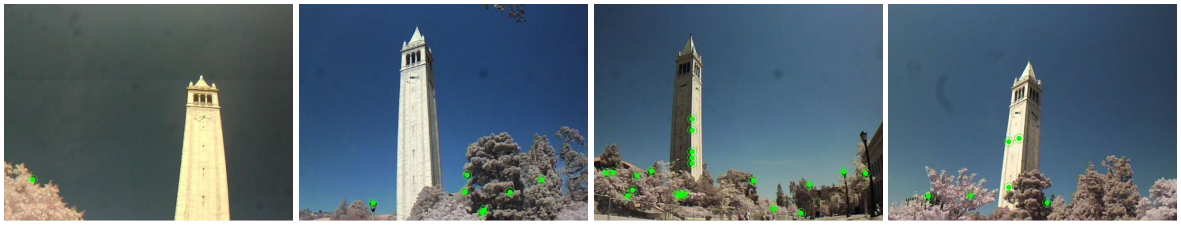
2.1. Failure of SfM on low-quality images

It was suggested by Turcot and Lowe [24] that the accuracy of object recognition in large image databases can be improved by suppressing uninformative visual words that typically represent irrelevant image background. In [24], SfM techniques were used to enforce pairwise epipolar constraints of rigid objects. The authors argued that, between

[1] In the literature, more sophisticated classifiers such as SVMs have also been used. Nevertheless, this is not the focus of the paper.



(a) Original SURF feature detection results.



(b) Informative features detected by SfM.

(c) Informative features selected by thresholded PCA based on the first two leading PVs.

(d) Informative features selected by Sparse PCA based on the first two leading PVs.

Figure 1. Comparison of informative feature selection on low-quality multiple-view images. A subset of 16 training images of a building (Campanile at UC Berkeley) in the BMW database [17] are used for training. For each image pair in SfM, SURF features are deemed informative if the consensus of the corresponding epipolar constraint exceeds 25% of the total feature pairs. For thresholded PCA, we manually assign small-valued entries to zero in PVs in attempt to achieve the same sparsity as Sparse PCA. The best results to identify informative features on the Campanile are given by Sparse PCA.

a pair of images that render the same object in space, uninformative features can be easily pruned out as outliers w.r.t. a dominant epipolar constraint by RANSAC. Along similar lines, Philbin *et al*. [20] introduced a Geometric Latent Dirichlet Allocation model for constructing image adjacency graphs. Subsequently, rich latent topic models were built from the adjacency graphs with the identity and locations of visual words specific to the objects, thereby rejecting uninformative visual words. Knopp *et al*. [11] augmented query images with rough geolocation information combined with wide-baseline feature matching to detect and suppress uninformative features before invoking vocabulary tree based object recognition.

All these methods rely on the accuracy of wide-baseline feature matching to establish pairwise epipolar geometry. However, they tend to fail when the quality of the images in the database is very poor, as is the case with images captured from mobile cellphones or distributed camera networks. Furthermore, man-made landmarks such as buildings often have repetitive texture and patterns that tend to confuse feature correspondence algorithms (*e.g*., Bundler [22]). Figure 1 (b) shows an example where SfM fails at determining the wide-baseline transformation across images of an object captured from multiple vantage points. More examples can be found in Figure 4 later.

3. Identifying Informative Features

Classical PCA is a well established tool for the analysis of high-dimensional data. For a data matrix A, PCA computes the PCs via an eigenvalue decomposition of its empirical covariance matrix Σ. It has also been observed that in general the entries of the corresponding PVs are dense and nonzero. In certain applications, it is desirable to obtain PVs that can explain maximum variability in the data A using linear combinations of just a few nonzero variables, and hence improves interpretability of such data. It is with this motivation that Sparse PCA was developed [5, 30] and has proven to be a very useful tool for identifying focalized hidden information in data where the coordinate axes involved have physical interpretations.

In the BoW approach to object recognition, each coordinate axis in the visual histogram corresponds to a particular visual word in the vocabulary tree. We contend that the visual words that explain maximum variability in data corresponding to each object category can be regarded as informative features for object recognition. In order to use Sparse PCA to identify these visual words, an empirical covariance matrix must first be computed for each object category in the database.

Let us consider m available training images of an object category. Using the constructed vocabulary tree learned from all the categories, the SURF descriptors in each image are converted into a visual histogram $\boldsymbol{y} \in \mathbb{R}^n$. The m vectors $\{\boldsymbol{y}_j\}$ are then normalized to have unit length and centered, and grouped into a *data matrix*: $A = [\tilde{\boldsymbol{y}}_1, \tilde{\boldsymbol{y}}_2, \cdots, \tilde{\boldsymbol{y}}_m] \in \mathbb{R}^{n \times m}$. The empirical covariance matrix is then computed from this data matrix as $\Sigma_A = \frac{1}{m} AA^T$.

Sparse PCA that computes the first sparse eigenvector of Σ_A optimizes the following objective [30]:

$$\boldsymbol{x}_s = \arg\max \boldsymbol{x}^T \Sigma_A \boldsymbol{x} \quad \text{subj. to} \quad \|\boldsymbol{x}\|_2 = 1, \|\boldsymbol{x}\|_1 \leq k. \tag{3}$$

We denote the indices of the non-zero coefficients in $\boldsymbol{x}_s$ by $\mathcal{I}$ (*i.e.*, the nonzero support of $\boldsymbol{x}_s$). These indices correspond to the visual words that explain maximum variability in A, and are subsequently used in the object recognition process (explained in Section 6).

In practice, it is common that the leading first sparse PV may not be sufficient for obtaining a variable support, and it is desirable to further estimate a few subsequent sparse PVs as well. In optimization, it is a common practice to estimate succeeding eigenvectors by sequentially deflating the covariance matrix with the preceding ones. Several techniques have been explored for reliably deflating a covariance matrix for Sparse PCA [16]. We adopt a simple technique called Hotelling's deflation that eliminates the influence of the first sparse PV to obtain a deflated covariance matrix Σ'_A as follows:

$$\Sigma'_A = \Sigma_A - (\boldsymbol{x}_s{}^T \Sigma_A \boldsymbol{x}_s) \boldsymbol{x}_s \boldsymbol{x}_s{}^T. \tag{4}$$

Then, the second sparse eigenvector $\boldsymbol{x}'_s$ of Σ_A becomes the leading sparse eigenvector of Σ'_A, and can be estimated again by Sparse PCA (3). In our experiment, we observe that the first two sparse PVs are sufficient for selecting informative features that lie on the foreground objects in the BMW database (as shown in Figure 1 and 4). Finally, If we denote the indices of the non-zeros in the second PV $\boldsymbol{x}'_s$ as $\mathcal{I}'$, then the union $\mathcal{I} \cup \mathcal{I}'$ provides the support corresponding to the informative features of a particular category.

For pedagogical purposes, we also compare the variable selection performance of thresholded PCA in Figures 1 and 4. To obtain a sparsified PCA support set, we perform PCA on the same covariance matrix Σ_A and pick the top k indices of the corresponding first and second PVs with highest absolute value as the informative features. Here, k is chosen as the same cardinality of the corresponding Sparse PVs for the same category. The examples clearly show that majority of the selected features do not represent the foreground objects.

4. Speeding up Sparse PCA using ALM

Sparse PCA has been an active research topic for over a decade. Notable approaches include SCoTLASS [10], SLRA [29], and SPCA [30], all of which aim at finding modified PVs with sparse entries. However, one drawback of all the above algorithms is that the formulation requires solving nonconvex objective functions. Recently, d'Aspermont *et al*. [5] derived an ℓ_1-norm based semidefinite relaxation for Sparse PCA called DSPCA, and it is currently the most widely known convex formulation of the problem. This algorithm, however, has a slow convergence rate that is a major bottleneck when analyzing high dimensional data. Augmented Lagrange multiplier (ALM) based algorithms have recently gained a lot of popularity due to their rapid convergence and speed in ℓ_1-minimization [26] and Robust PCA [14] problems. These have motivated us to develop a new algorithm for solving the semidefinite relaxation form of Sparse PCA using ALM.

We begin by showing Sparse PCA can be converted to a SDP [5]. Given an empirical covariance matrix $\Sigma \in \mathbb{S}^n$, with n representing the dimensionality of the data, Sparse PCA solves the following objective:

$$\max_{\|\boldsymbol{x}\|_2 \leq 1} \boldsymbol{x}^T \Sigma \boldsymbol{x} - \rho \|\boldsymbol{x}\|_0, \tag{5}$$

where $\rho > 0$ is a scalar parameter controlling the sparsity in $\boldsymbol{x}$. By following the ℓ_1-norm relaxation and lifting procedure for semidefinite relaxation, and dropping a nonconvex rank constraint, we can rewrite (5) as [5]:

$$\max_X \mathbf{Tr}(\Sigma X) - \rho \|X\|_1 \ : \ \mathbf{Tr}(X) = 1, X \succeq 0, \tag{6}$$

where $X = \boldsymbol{x}\boldsymbol{x}^T$ is a matrix variable. Duality allows us to rewrite this problem as a SDP:

[2] In this paper, $\|X\|_1$ represents the entrywise norm: $\mathbf{1}^T |X| \mathbf{1}$.

$$\min_U \ \lambda_{\max}(\Sigma + U) \ : \ -\rho \leq U_{ij} \leq \rho. \quad (7)$$

As presented in [5], assuming Σ is fixed and given, the maximum eigenvalue function $\lambda_{max}(\cdot)$ can be approximated by a smooth, uniform objective (*i.e.*, with Lipschitz continuous gradient):

$$f_\mu(U) = \mu \log(\mathbf{Tr} \exp((\Sigma + U)/\mu)) - \mu \log(n), \quad (8)$$
$$\nabla f_\mu(U) = \exp((\Sigma + U)/\mu)/\mathbf{Tr}(\exp((\Sigma + U)/\mu)), \quad (9)$$

where $\mu = \epsilon/2\log(n)$ produces an ϵ-approximate solution. With this approximation, (7) can be rewritten as:

$$\min_U \ f_\mu(U) \ : \ -\rho \leq U_{ij} \leq \rho. \quad (10)$$

Based on the above SDP formulation, next we consider speeding up Sparse PCA via an ALM approach [2]. The basic idea is to eliminate the constraints and add to the cost function a penalty term that prescribes a high cost to infeasible points. This augmented cost function is called the *augmented Lagrangian function*. In our case, the box constrained convex problem of (10) can be written in an unconstrained form as:

$$F(U, Y) \doteq \min_U \{f_\mu(U) + \sum_{1 \leq i,j \leq n} P(U_{ij}, Y_{ij}, c)\}, \quad (11)$$

where $Y_{ij}, 1 \leq i, j \leq n$ represents the Lagrange variable, c determines the severity of the penalty, and

$$P(u, y, c) = \begin{cases} y(u - \rho) + \frac{c}{2}(u - \rho)^2 & \text{if } \rho - \frac{y}{c} \leq u, \\ y(u + \rho) + \frac{c}{2}(u + \rho)^2 & \text{if } -\rho - \frac{y}{c} \geq u, \\ \frac{y^2}{2c} & \text{otherwise.} \end{cases} \quad (12)$$

The algorithm for Sparse PCA using ALM (SPCA-ALM) is presented in Algorithm 1. Note that in each iteration of the outer loop of the algorithm, we need to solve the unconstrained minimization problem in (11), which has no closed-form solution. Thus, we employ Nesterov's first order gradient technique [18]. Once this augmented Lagrangian function is minimized, the Lagrange multipliers Y will be updated using the rule:

$$Y_{ij}^{k+1} = \begin{cases} Y_{ij}^k + c^k(U_{ij}^k - \rho) & \text{if } Y_{ij}^k + c^k(U_{ij}^k - \rho) > 0, \\ Y_{ij}^k + c^k(U_{ij}^k + \rho) & \text{if } Y_{ij}^k + c^k(U_{ij}^k + \rho) < 0, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

After the algorithm converges, the primal variable is given by the gradient in (9), *i.e.*, $X^k = \nabla f_\mu(U^k)$. Then the sparse principal component is recovered as the leading eigenvector of X^k.

4.1. Performance

We have evaluated our SPCA-ALM algorithm by comparing its performance against the DSPCA solver [5]. Both

Algorithm 1: SPCA-ALM

Input: Covariance Σ and $\rho > 0$.

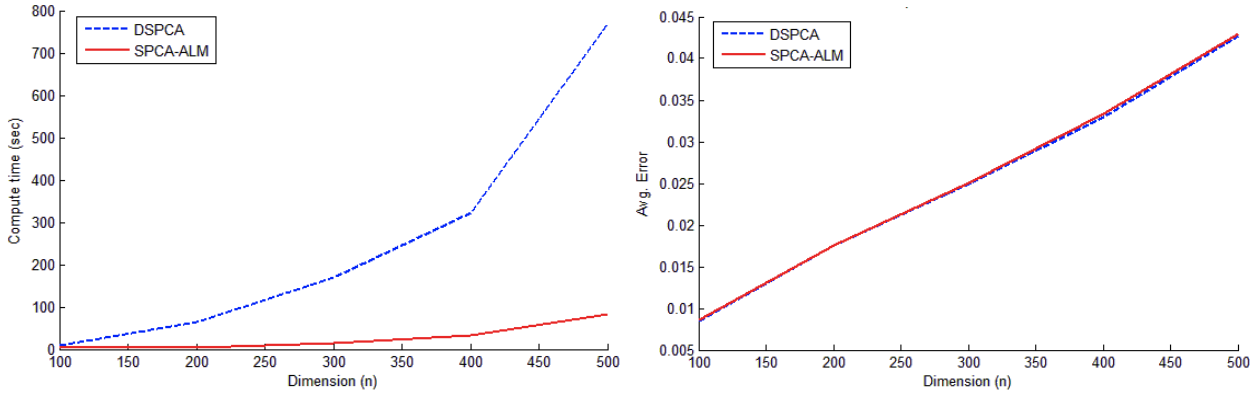
$U^1 \leftarrow \mathbf{0}$, $Y^1 \leftarrow \mathbf{0}$, $X^1 \leftarrow \mathbf{0}$, $c^1 \leftarrow 1$.
while not converged (k=1,2,3,...) **do**
  $t^1 \leftarrow 1$, $V^1 \leftarrow U^k$, $W^0 \leftarrow U^k$, $Z \leftarrow rand(n, n)$.
  $\alpha^0 \leftarrow \frac{\|V^1 - Z\|_F}{\|\nabla F(V^1, Y^k) - \nabla F(Z, Y^k)\|_F}$.
  while not converged (l=1,2,3,...) **do**
    Find smallest $i \geq 0$ for which
    $F(V^l, Y^k) - F(V^l - \frac{\alpha^{l-1}}{2^i}\nabla F(V^l, Y^k), Y^k) \geq \frac{\alpha^{l-1}}{2^{i+1}}\|\nabla F(V^l, Y^k)\|_F$.
    $\alpha^l \leftarrow 2^{-i}\alpha^{l-1}$, $W^l \leftarrow V^l - \alpha^l \nabla F(V^l, Y^k)$.
    $t^{l+1} \leftarrow (1 + \sqrt{4t^{l^2} + 1})/2$.
    $V^{l+1} \leftarrow W^l + \frac{t^l - 1}{t^{l+1}}(W^l - W^{l-1})$.
  end while
  $U^{k+1} \leftarrow W^l$
  Update Y^{k+1} using the update rule (13).
  $X^{k+1} \leftarrow \nabla f_\mu(U^{k+1})$.
  $c^{k+1} \leftarrow 2^k$.
end while

Output: Sparse principal vector, $x_s \leftarrow$ leading eigenvector of X^k.

algorithms have been implemented in MATLAB and benchmarked on a 2.6 GHz Intel processor with 4 GB memory. We generate synthetic data of varying dimensionality as follows. First, in the n-dimensional vector space, 10% of its indices are selected as nonzero support. Next, the values of the nonzero coefficients are drawn from an independent and identically distributed Gaussian $\boldsymbol{x}_0(i) \sim N(0, 200)$. Finally, random noise $\epsilon \sim N(0, 1)$ is added to x_0 to form a noisy version of the empirical covariance matrix, $\Sigma = (\boldsymbol{x}_0 + \epsilon\mathbf{1})(\boldsymbol{x}_0 + \epsilon\mathbf{1})^T$. This covariance matrix, along with an optimal choice of the parameter ρ to encourage sparsity, is provided to both the SPCA-ALM and DSPCA algorithms. The process repeats 10 times for each problem dimension n, while n varies from 100 to 500 and the average speed and precision are computed for each n. Figure 2(a) compares the speed of the two algorithms, while Figure 2(b) compares the estimation error of the first estimated sparse principal vector. The simulation shows SPCA-ALM converges much faster than DSPCA (for example, at $n = 500$, SPCA-ALM is about 10 times faster), while maintaining approximately the same reconstruction accuracy.

5. Variable Elimination via SAFE

In this section, we further examine a dimensionality reduction technique as a preprocessing step to speed up Sparse PCA. Particularly in object recognition, the covariance matrix Σ often can be of high dimension (*e.g.*, 1000 and higher). Directly calling SPCA-ALM may still be



(a) Speed vs Data Dimension (b) Estimation Error vs Data Dimension

Figure 2. A comparison of SPCA-ALM and DSPCA using simulated data.

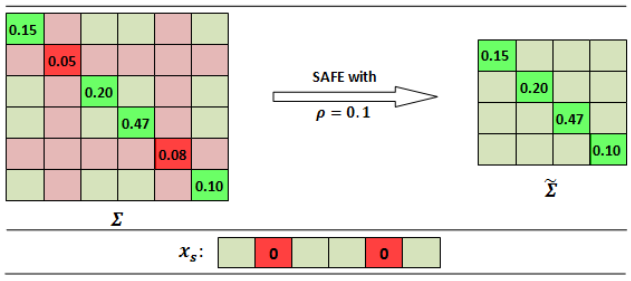


Figure 3. SAFE feature elimination process. **Top:** The red rows and columns of a sample covariance matrix Σ are eliminated to form new covariance matrix $\tilde{\Sigma}$, as the corresponding variances are less than chosen $\rho = 0.1$. **Bottom:** The entries of the corresponding indices are subsequently zeroed out in x_s.

very time consuming. To mitigate this problem, we invoke a feature elimination method presented in [6, 28], called SAFE. The method allows to quickly eliminate variables in problems involving a convex loss function and a ℓ_1-norm penalty, thereby leading to substantial reduction in the number of variables prior to running optimization. The following theorem [6, 28] states the SAFE method applied to Sparse PCA. An illustration of this process is shown in Figure 3.

Theorem 1 (SAFE Variable Elimination for Sparse PCA)**.** *Given a covariance matrix Σ, denote σ_k as its kth diagonal entry. For the Sparse PCA problem (5), if $\rho > \sigma_k$, then the kth element of the solution $\boldsymbol{x}_s$ will not be in the sparse support. Hence, the kth row and column of Σ can be removed from the optimization.*

Therefore, for a predefined choice of ρ, we first obtain a reduced covariance matrix by eliminating all the rows and columns corresponding to those variables with sample variance less than ρ. The number of variables thus eliminated is a conservative lower bound on the total number of zero-weight variables in the final solution of Sparse PCA. In our experiments, we typically can eliminate about 90% of the variables using SAFE without sacrificing the accuracy of preserving important informative features.

6. Experiment

In order to test the effectiveness of suppressing uninformative features for the task of object recognition, we have evaluated the performance of our method on the Berkeley Multiview Wireless (BMW) database [17]. The database consists of multiple-view images of 20 landmark buildings on the Berkeley campus. For each building, wide-baseline images were captured from 16 different vantage points. Further, at each vantage point, 5 short-baseline images were taken (by five camera sensors #0 – #4 simultaneously), thereby summing to 80 images per category. All images are 640×480 RGB color images. It is important to note that the image quality in this database is considerably lower than many existing high-resolution databases, which is intended to reproduce realistic imaging conditions for mobile camera and surveillance applications. Further, it is noticeable that some images are slightly out of focus and in some cases, even corrupted by dust residual on the camera lenses.

We divide the database into a training set and a testing set. The vantage points of each object are named numerically from 0 to 15. All these 16 images of each category captured from camera #2 are designated as the training set, and the remaining images are assigned to the testing set. Thus, there are 16 training images and 64 testing images for each category. We extract SURF features in each of the training images and construct a vocabulary tree with 1000 leaf nodes.

6.1. Results

We first evaluate the recognition accuracy of the classifier (2) without suppressing any features from the training and testing sets to obtain a baseline performance. The results of this experiment are presented in Table 1. For the 20 object categories tested, the average baseline recognition rate is around 80%.

Next, for each object category i, we obtain its informative feature set $\mathcal{I}_i$ by determining the indices of the non-zero variables in the first and second sparse PVs. These are estimated by running Sparse PCA on the covariance matrix corresponding to the training histogram vectors in ith category. We then form the total support set $\mathcal{I}_{\text{SPCA}}$ for the entire database by taking the union of the individual visual support sets for all the 20 object categories, *i.e.*,

$$\mathcal{I}_{\text{SPCA}} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \cdots \cup \mathcal{I}_{20}.$$

In our experiments, we have set the sparsity controlling parameter ρ to 0.002 for all the categories. With this choice of ρ, at roughly 33 variables per category, our total support set $\mathcal{I}_{\text{SPCA}}$ identifies 405 informative features (some informative features overlap between classes), thereby rejecting a fraction of $\frac{3}{5}$ of the visual words from the 1000-D vocabulary. With this subset of visual words, we evaluate the recognition accuracy of (2) again. The results are also presented in Table 1. As one can see, for most of the categories, there is a significant improvement in the recognition accuracy, leading to the average recognition rate at 85%, 5% higher than the baseline.

For completeness, Table 1 also shows the number of selected features and the recognition rates for the SfM approach. For a large number of the object categories, the SfM method does not seem to work well, as few of the SURF features are correctly selected as foreground features. We have tested the recognition accuracy of these visual words on the database as well, and the average rate is 78%, even lower than that of the baseline performance. Finally, some visual comparisons between the results from Sparse PCA and SfM are presented in Figure 4.

7. Conclusion and Discussion

We have presented a novel and effective solution to select informative features for object recognition by Sparse PCA. For applications that involve low-quality mobile cameras or surveillance camera networks, existing SfM solutions to detect and suppress uninformative features tend to fail. We have shown that Sparse PCA can successfully identify important visual features that explain maximum variability in the visual histogram vectors. For our database, these features correspond to those visual words that most often represent the appearance of foreground objects. To further speed up the execution of Sparse PCA, we have developed an improved numerical algorithm, namely, SPCA-ALM. The new algorithm has proved significantly faster than the other convex semidefinite programming solutions. Using a public multiple-view image database, our experiment shows the estimated informative features improve the overall recognition rate by 5% compared to the baseline solution, and by 7% compared to the SfM solution.

For future work, we believe the two existing approaches, namely, Sparse PCA and SfM, are complementary under more general object recognition settings. We would like to focus on further combining our batch numerical technique within a geometric RANSAC scheme to robustly detect informative features in both low-quality and high-quality image databases, which may lead to further improvement of the performance.

Table 1. Recognition rates and number of selected informative features for the 20 object classes in alphabetical order [17]. The best rates are in bold face. The categories in which SfM failed have zero feature selected.

Cat.	Baseline Rate(%)	SPCA Rate(%)	SPCA # Feat	SfM Rate(%)	SfM # Feat
1	**98.61**	94.44	35	83.33	0
2	90.27	**91.66**	23	90.27	35
3	56.94	**66.66**	15	58.33	0
4	70.83	**81.94**	12	65.27	30
5	77.77	**91.66**	56	81.94	0
6	**95.83**	88.88	23	87.50	0
7	79.16	**93.05**	34	86.11	0
8	77.77	**91.66**	30	72.22	0
9	56.94	**73.61**	45	63.88	11
10	51.38	**65.27**	9	61.11	0
11	**83.33**	76.38	76	69.44	13
12	81.94	**83.33**	28	70.83	0
13	62.50	**72.22**	43	52.77	0
14	**98.61**	93.05	20	90.27	37
15	69.44	**80.55**	36	75.00	0
16	58.33	79.16	53	**80.55**	66
17	**100.00**	90.27	17	84.72	0
18	98.61	93.05	45	**100.00**	56
19	**97.22**	83.33	24	86.11	0
20	98.61	**100**	46	95.83	0
Avg.	80.02	**84.51**	33	77.77	12

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. *CVIU*, 110(3):346–359, 2008. 2

[2] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2008. 2, 5

[3] Z. Cheng, D. Devarajan, and R. Radke. Determining vision graphs for distributed camera networks using feature digests. *EURASIP J. Adv. in Sig. Proc.*, pages 1–11, 2007. 1

[4] C. Christoudias, R. Urtasun, and T. Darrell. Unsupervised feature selection via distributed coding for multi-view object recognition. In *CVPR*, 2008. 1

[5] A. d'Aspremont, L. El Ghaoui, M. Jordan, and G. Lanckriet. A direct formulation for sparse pca using semidefinite programming. *SIAM Rev.*, 2007. 2, 4, 5

[6] L. El Ghaoui, V. Viallon, and T. Rabbani. Safe feature elimination in sparse supervised learning. Technical Report UCB/EECS-2010-126, UC Berkeley, 2010. 6

[7] L. Fei-Fei. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, 2005. 1

[8] V. Ferrari, T. Tuytelaars, and L. Van Gool. Integrating multiple model views for object recognition. In *CVPR*, 2004. 1

Figure 4. **Top:** Images of 6 objects in the BMW database with superimposed SURF features; **Middle-top:** Informative features detected by the SfM approach; **Middle-bottom:** Informative features detected by thresholded PCA (first two leading PVs); **Bottom:** Informative features detected by Sparse PCA (first two leading sparse PVs).

[9] Y. Jiang, C. Ngo, and J. Yang. Towards optimal bag-of-features for object categorization and semantic video retrieval. In *ACM Int. Conf. on Image and Video Retrieval*, 2007. 1

[10] I. Jolliffe, N. Trendafilov, and M. Uddin. A modified principal component technique based on the lasso. *JCGS*, 2003. 4

[11] J. Knopp, J. Sivic, and T. Pajdla. Avoiding confusing features in place recognition. In *ECCV*, 2010. 3

[12] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006. 1

[13] J. Lee. Libpmk: A pyramid match toolkit. Technical Report MIT-CSAIL-TR-2008-017, MIT, 2008. 2

[14] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical Report UILU-ENG-09-2215, UIUC, 2009. 4

[15] D. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999. 2

[16] L. Mackey. Deflation methods for Sparse PCA. In *NIPS*, 2009. 4

[17] N. Naikal, A. Yang, and S. Sastry. Towards an efficient distributed object recognition system in wireless smart camera networks. In *Information Fusion*, 2010. 1, 3, 6, 7

[18] Y. Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). *Soviet Mathematics Doklady*, 1983. 5

[19] D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In *CVPR*, 2006. 1, 2

[20] J. Philbin and J. S. A. Zisserman. Geometric latent dirichlet allocation on a matching graph forlarge-scale image datasets. *IJCV*, 2010. 3

[21] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, 2003. 1

[22] N. Snavely, S. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *IJCV*, 2007. 2, 3

[23] A. Thomas, V. Ferrari, B. Leibe, T. Tuytelaars, B. Schiele, and L. Van Gool. Towards multi-view object class detection. In *CVPR*, 2006. 1

[24] P. Turcot and D. Lowe. Better matching with fewer features: The selection of useful features in large database recognition problems. In *ICCV Workshop on Emergent Issues in Large Amounts of Visual Data*, 2009. 1, 2

[25] A. Yang, S. Maji, C. Christoudias, T. Darrell, J. Malik, and S. Sastry. Multiple-view object recognition in band-limited distributed camera networks. In *ICDSC*, 2009. 1

[26] A. Yang, Z. Zhou, Y. Ma, and S. Sastry. Fast ℓ_1-minimization algorithms and an application in robust face recognition: A review. In *ICIP*, 2010. 2, 4

[27] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *IJCV*, 2007. 1

[28] Y. Zhang, A. d'Aspremont, and L. El Ghaoui. Sparse PCA: convex relaxations, algorithms and applications. Technical Report arXiv:1011.3781, arXiv, 2010. 6

[29] Z. Zhang, H. Zha, and H. Simon. Low-rank approximations with sparse factors I: Basic algorithms and error analysis. *SIAM J. Matrix Analysis Applications*, 2002. 4

[30] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *JCGS*, 2006. 2, 4